

EXAMINER'S AMENDMENT

1. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

Authorization for this examiner's amendment was given in a telephone interview with Tawni Wilhelm (Reg. No. 47,456) on November 16, 2009.

Please amend the claims as follows:

1. (Currently Amended) A system that facilitates incremental web crawls comprising the following components stored in computer memory and executable by a processor:

a selecting component that selects a first chunk from a group that includes at least one index chunk that stores information associated with an index of items, at least one rank chunk that stores at least one static rank associated with the at least one index chunk, at least one content chunk that stores cached copies of contents of pages crawled, at least one re-crawl chunk that stores a list of Uniform Resource Locators to be re-crawled, and at least one webmap chunk that stores at least a portion of a web map used for calculating the at least one static rank;

a chunk map that stores properties associated with the first chunk and that is employed to determine the first chunk based on the stored properties, and wherein the stored properties include average time between change and average importance and are shared by all items in the first chunk;

a parsing component that parses the first chunk for Uniform Resource Locators;
a crawling component that re-crawls the Uniform Resource Locators;
a receiving component that

(1) receives one or more documents as a result of re-crawling the uniform resource locators,

(2) stores a document from the one or more documents in an appropriate chunk when it is determined that the document belongs to the appropriate chunk, and

(3) forms a second chunk separate from the first chunk based, at least in part, upon the re-crawled uniform resource locators when it is determined that the document does not belong to any chunk from the group, and storing the document in the second chunk

~~an indexer that places items with similar properties into respective chunks, wherein the properties are shared by all the items within a respective chunk, wherein the items are the results returned by a web crawl, and wherein the respective chunks include at least one rank chunk and at least one webmap chunk; and~~

~~a chunk map that stores at least some of the properties associated with each respective chunk, wherein the properties include average time between change and average importance of documents in each respective chunk, wherein the chunk map is employed to facilitate an incremental web re-crawl, and wherein the properties of each respective chunk stored in the chunk map are utilized to determine a re-crawl of all the items in that respective chunk.~~

2. (Previously Presented) The system of claim 1, wherein the items comprise information associated with a Uniform Resource Locator.

3. (Previously Presented) The system of claim 1, wherein the items comprise at least one of an HTML file, a PDF file, a PS file, a PPT file, an XLS file and a DOC file.

4. (Previously Presented) The system of claim 1, wherein the items are received from a crawler, and wherein the crawler is responsible for a specific set of Uniform Resource Locators.

5. (Original) The system of claim 1, further comprising a master control process that can modify the chunk map to facilitate load balancing amongst a plurality of crawlers.

6. (Original) The system of claim 1, further comprising a master control process that serves as an interface between a crawler and a re-crawl controller.

7. (Previously Presented) The system of claim 6, wherein the master control process maintains a known chunks table that stores information for components of the system.

8. (Original) The system of claim 6, wherein the master control process exposes an interface for communication with a component of the system.

9. (Original) The system of claim 8, wherein the interface returns a list of chunks the component should have and where to get the chunks.

10. (Currently Amended) The system of claim 8, wherein the interface returns a list of ~~the~~ chunks that should be actively served by the component.

11. (Original) The system of claim 8, wherein the interface returns a range of chunk identifiers to use in building a new chunk by the component.

12. (Original) The system of claim 8, wherein the interface causes an old chunk to be retired by the system.

13. (Original) The system of claim 6, wherein the master control process facilitates movement of chunks from one component to another component.

14. (Previously Presented) The system of claim 13, wherein movement of chunks is based, at least in part, upon at least one of rebalancing index servers after one goes down, re-crawling pages previously crawled, and restoring a state of the crawler after it has crashed.

15. (Currently Amended) The system of claim 1, ~~further comprising a re-crawl component that wherein the crawling component employs the chunk map to determine which chunks, if any, to re-crawl at a particular time.~~

16. (Cancelled)

17. (Original) The system of claim 1, further comprising an index chunk that stores information associated with an index of at least some of the items.

18. (Cancelled)

19. (Currently Amended) Computer-readable storage media having computer-useable instructions embodied thereon for performing a method of A method of performing document re-crawl, the method comprising:

selecting a first chunk from a group that includes at least one index chunk that stores information associated with an index of items, at least one rank chunk that stores at least one static rank associated with the at least one index chunk, at least one content chunk that stores cached copies of contents of pages crawled, at least one re-crawl chunk that stores a list of Uniform Resource Locators to be re-crawled, and at least one webmap chunk that stores at least a portion of a web map used for calculating the at least one static rank, wherein a chunk map that stores properties associated with the first chunk is employed to determine the first chunk based on the stored properties,

and wherein the stored properties include average time between change and average importance and are shared by all items in the first chunk;

parsing the first chunk for Uniform Resource Locators;

re-crawling the Uniform Resource Locators;

receiving one or more documents as a result of re-crawling the uniform resource locators;

storing a document from the one or more documents in an appropriate chunk when it is determined that the document belongs to the appropriate chunk; and,

forming a second chunk separate from the first chunk based, at least in part, upon the re-crawled uniform resource locators when it is determined that the document does not belong to any chunk from the group, and storing the document in the second chunk.

20. (Currently Amended) The ~~method~~ media of claim 19, wherein the method further comprising comprises at least one of the following acts:

determining whether any chunks are to be retired;

moving the first chunk; and

destroying the first chunk.

21. (Canceled)

22. (Currently Amended) A method of performing document re-crawl comprising:

selecting, utilizing a first computing process, a first chunk from a group that includes at least one index chunk that stores information associated with an index of items, at least one rank chunk that stores at least one static rank associated with the at least one index chunk, at least one content chunk that stores cached copies of contents of pages crawled, at least one re-crawl chunk that stores a list of Uniform Resource Locators to be re-crawled, and at least one webmap chunk that stores at least a portion of a web map used for calculating the at least one static rank, wherein a chunk map that stores properties associated with the first chunk is employed to determine the first chunk based on the stored properties, and wherein the stored properties include average time between change and average importance and are shared by all items in the first chunk;

parsing, utilizing a second computing process, the first chunk for Uniform Resource Locators;

re-crawling, utilizing a third computing process, the Uniform Resource Locators; receiving one or more documents as a result of re-crawling the uniform resource locators;

storing a document from the one or more documents in an appropriate chunk when it is determined that the document belongs to the appropriate chunk; and,
forming, utilizing a fourth computing process, a second chunk separate from the first chunk based, at least in part, upon the re-crawled uniform resource locators when it

is determined that the document does not belong to any chunk from the group, and
storing the document in the second chunk,

wherein the first, second, third and fourth computing processes are performed by
one or more computing devices

~~parsing a first chunk for Uniform Resource Locators, wherein the Uniform~~
~~Resource Locators are stored as a result of one or more web crawls;~~
~~accessing a chunk map comprising properties associated with one or more~~
~~chunks that include the first chunk, wherein the stored properties are shared by all the~~
~~items in the first chunk, and wherein the properties include average time between~~
~~change and average importance of documents in the first chunk; and,~~

~~periodically determining, based on the properties of each of the one or more~~
~~chunks in the chunk map, whether to re-crawl the items in the first chunk.~~

23. (Currently Amended) The method of claim 22, further comprising
periodically determining whether to re-crawl the Uniform Resource Locators based, at
least in part, upon at least one of the average time between change and the average
importance of documents comprising a particular chunk wherein the periodic
determination is based, at least in part, upon at least one of average time between
change and average importance of documents comprising a particular chunk.

24-28. (Canceled)

REASONS FOR ALLOWANCE

2. The following is an examiner's statement of reasons for allowance: Independent claim 1 is drawn to a system that facilitates incremental web crawls, independent claim 19 to computer-readable storage media containing instructions for performing a method of document re-crawl, and independent claim 22 to a method of performing document re-crawl. The system, computer-readable storage media and method, among other things, all utilize a plurality of different chunks and select a chunk from the plurality of chunks and re-crawl the URLs in the selected chunk. The chunks include at least one index chunk, at least one rank chunk, at least one content chunk, at least one re-crawl chunk and at least one webmap chunk.

The prior art references made of record to not teach, either together or separately, a web crawler that selects a first chunk from the group including all of the chunks mentioned above in order to re-crawl the URLs within the selected chunk. Nor does the prior art of record teach forming a second chunk separate from the first chunk based upon the re-crawled URLs when it is determined that the document does not belong to any chunk from the group.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Scott M. Sciacca whose telephone number is (571) 270-1919. The examiner can normally be reached on Monday thru Friday, 7:30 A.M. - 5:00 P.M. EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Jeff Pwu can be reached on (571) 272-6798. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Scott M. Sciacca/
Examiner, Art Unit 2446

/Benjamin R Bruckart/
Primary Examiner, Art Unit 2446